

SUBCONVEXITY AND EQUIDISTRIBUTION OF HEEGNER POINTS IN THE LEVEL ASPECT

SHENG-CHI LIU, RIAD MASRI, AND MATTHEW P. YOUNG

ABSTRACT. Let q be a prime and $-D < -4$ be an odd fundamental discriminant such that q splits in $\mathbb{Q}(\sqrt{-D})$. For f a weight zero Hecke-Maass newform of level q and Θ_χ the weight one theta series of level D corresponding to an ideal class group character χ of $\mathbb{Q}(\sqrt{-D})$, we establish a hybrid subconvexity bound for $L(f \times \Theta_\chi, s)$ at $s = 1/2$ when $q \asymp D^\eta$ for $0 < \eta < 1$. With this circle of ideas, we show that the Heegner points of level q and discriminant D become equidistributed, in a natural sense, as $q, D \rightarrow \infty$ for $q \leq D^{1/20-\varepsilon}$. Our approach to these problems is connected to estimating the L^2 -restriction norm of a Maass form of large level q when restricted to the collection of Heegner points. We furthermore establish bounds for quadratic twists of Hecke-Maass L -functions with simultaneously large level and large quadratic twist, and hybrid bounds for quadratic Dirichlet L -functions in certain ranges.

Compositio Mathematica (2013), to appear.

1. INTRODUCTION AND STATEMENT OF RESULTS

Let f be an arithmetically normalized Hecke-Maass newform of weight zero and prime level q with spectral parameter t_f . Let $K = \mathbb{Q}(\sqrt{-D})$ be an imaginary quadratic field of discriminant $-D < -4$, such that q splits in K , with ideal class group CL_K and class number $h(-D)$. Given an ideal class group character $\chi \in \widehat{\text{CL}}_K$, let Θ_χ be the weight one theta series of level D corresponding to χ .

Theorem 1.1. *With notation as above, we have*

$$\sum_{\chi \in \widehat{\text{CL}}_K} L(f \times \Theta_\chi, \tfrac{1}{2}) = \frac{3}{\pi} \frac{h(-D)^2}{\sqrt{D}} \frac{q^2}{q^2 - 1} L(\text{sym}^2 f, 1) + O_{t_f, \varepsilon}((qD)^\varepsilon \min(qD^{7/16}, q^{3/4}D^{1/4} + q^{1/4}D^{1/2})). \quad (1.1)$$

We have established an analogue of Theorem 1.1 for holomorphic forms f which will appear in a subsequent paper.

Recall the conductor of $L(f \times \Theta_\chi, s)$ at $s = 1/2$ is $Q = (qD)^2$, so the convexity bound is $L(f \times \Theta_\chi, \frac{1}{2}) \ll_{t_f, \varepsilon} Q^{\frac{1}{4} + \varepsilon}$. These central values are nonnegative, and dropping all but one term in Theorem 1.1 yields a subconvexity bound when $q \asymp D^\eta$ for $0 < \eta < 1$. There are a variety of cases to consider to state the best bound as a function of η ; for simplicity we

2010 *Mathematics Subject Classification.* 11M41.

Key words and phrases. subconvexity, equidistribution, Heegner points, L^4 norm.

R.M and M.Y. were supported by the National Science Foundation under agreement Nos. DMS-1162535 (R.M.) and DMS-0758235 (M.Y.). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

record what one obtains with the second bound in (1.1) which suffices for subconvexity for $0 < \eta < 1$.

Corollary 1.2. *For $\eta := \log(q)/\log(D)$ satisfying $0 < \eta < 1$, we have*

$$L(f \times \Theta_\chi, \frac{1}{2}) \ll_{t_f, \varepsilon} Q^{\frac{1}{4} + \varepsilon} \left(Q^{-\frac{1-\eta}{8(1+\eta)}} + Q^{-\frac{\eta}{8(1+\eta)}} \right). \quad (1.2)$$

We note that Theorem 1.1 also yields the following quantitative nonvanishing result.

Corollary 1.3. *For each $\varepsilon > 0$ there is an effective constant $c = c(\varepsilon, t_f) > 0$ such that whenever $q \leq cD^{1/16-\varepsilon}$, we have $L(f \times \Theta_\chi, 1/2) \neq 0$ for some $\chi \in \widehat{\text{CL}}_K$.*

The subconvexity problem for L -functions in the level aspect has been studied extensively in recent years. For example, let f and g be Hecke cusp forms (holomorphic or Maass) for GL_2 of levels M and N respectively. If M is fixed and N varies, subconvexity bounds of the form $L(f \times g, 1/2) \ll_f N^{\frac{1}{2}-\delta}$ for some absolute $\delta > 0$ were established in various settings by many authors [KMV] [Mi] [HaMi] [MV3]. On the other hand, it is also of interest to establish a subconvexity bound for $L(f \times g, 1/2)$ when *both* M and N are allowed to vary. A model result of this type was established by Michel and Ramakrishnan [MR], who considered the average of $L(f \times \Theta_\chi, 1/2)$ over holomorphic forms f . Their result also implies subconvexity for these L -functions in the wide range $0 < \eta < 1$ (with notation as in Corollary 1.2). They remark that this subconvexity result is intriguing because “such uniformity seems hard to achieve by purely analytic methods.” Here we average over χ (not f), though after some transformations we are led to averaging $L(g \times \chi_D, 1/2)$ over g , where g runs over level q Hecke-Maass forms and χ_D is a quadratic Dirichlet character. Feigon and Whitehouse [FW] generalized the work of [MR] to the number field setting, and Nelson [N2] has obtained results with Θ_χ replaced by more general holomorphic cusp forms. Using a different approach (a second moment), Holowinsky and Munshi [HoMu] have obtained subconvexity when $M \asymp N^\eta$ for $\eta > 0$ in some fixed range; their work does not require the central values to be nonnegative. It is also natural to consider the average over f instead of θ_χ in (1.1); by this approach, Holowinsky and Templier [HoTe] have recently shown a hybrid subconvexity bound in the range $0 < \eta < 1$.

The L -functions in Theorem 1.1 arise naturally in various arithmetic problems related to the equidistribution of Heegner points (see e.g. [D] [Z1] [Z2] [Mi] [HaMi] [MV1] [MV2]). We will use some ideas involved in the proof of Theorem 1.1 to prove a “sparse” equidistribution theorem for Galois orbits of Heegner points in which both q and D are varying; this is different from the more familiar case as in [Mi] [HaMi] where for q fixed and D varying the sparsity comes from suborbits of the full Galois orbit. See Section 7 for a description of the problem and Theorem 7.1 for the precise result. For brevity, we state here the following special consequence.

Let H be the Hilbert class field of K and $G := \text{Gal}(H/K) \cong \text{CL}_K$. The set $\Lambda_D(q)$ of Heegner points of discriminant $-D$ on the modular curve $X_0(q)$ splits into two simple, transitive G -orbits which are permuted by the Fricke involution w_q which acts on weight 0 forms f (invariant under $\Gamma_0(q)$) by $(w_q f)(z) = f(-1/qz)$; for these facts, see pp. 235-236 of [GZ]. For $\tau \in \Lambda_D(q)$, consider the Galois orbit $G\tau = \{\tau^\sigma : \sigma \in G\}$. The (open) modular curve $Y_0(q) = \Gamma_0(q) \backslash \mathbb{H}$ can be written as

$$Y_0(q) = \bigcup_{\omega_q \in \Gamma_0(1)/\Gamma_0(q)} \omega_q^{-1} Y_0(1).$$

For any given ω_q , we are interested in an asymptotic count for the number

$$N_{G,D,\omega_q} := \#(G\tau \cap \omega_q^{-1}Y_0(1))$$

of Heegner points in $G\tau$ which lie in $\omega_q^{-1}Y_0(1)$ as $q, D \rightarrow \infty$. Here we have in mind an analogy with counting primes (say) in an arithmetic progression where the choice of ω_q corresponds to the choice of residue class modulo q in which one is counting primes. Note that by the $SL_2(\mathbb{R})$ -invariance of the hyperbolic measure we have

$$\frac{\text{vol}(\omega_q^{-1}Y_0(1))}{\text{vol}(Y_0(q))} = \frac{1}{q+1},$$

and thus the volume of $\omega_q^{-1}Y_0(1)$ is becoming very small compared to the total volume of $Y_0(q)$ as $q \rightarrow \infty$.

Theorem 1.4. *With notation as above we have, uniformly in ω_q ,*

$$N_{G,D,\omega_q} \sim \frac{h(-D)}{q+1} \quad (1.3)$$

as $q, D \rightarrow \infty$ with the restriction $q \leq D^{1/20-\varepsilon}$. Furthermore, for $D^{1/20-\varepsilon} \leq q \leq D^{1/4-\varepsilon}$, we have

$$N_{G,D,\omega_q} \ll q^{1/4} D^{7/16+\varepsilon}. \quad (1.4)$$

In particular, we obtain: given q and ω_q , there exists a Heegner point of discriminant D in $\omega_q^{-1}Y_0(1)$ with some $D \ll_H q^H$ for any $H > 20$. This is a natural analog of Linnik's theorem on the first prime in an arithmetic progression: given q and any allowable residue class $a \pmod{q}$, there exists a prime $p \equiv a \pmod{q}$ with $p \ll q^L$ where L is the famous Linnik constant (see Chapter 18 of [IK]).

The upper bound in (1.4) shows that the Heegner points cannot cluster too much into one translate $\omega_q^{-1}Y_0(1)$.

Our work here has some connections with L^2 restriction norms of automorphic forms. The formula (4.1) below relates the average of L -values appearing in Theorem 1.1 to the sum over Heegner points of the modulus squared of a level q Maass form. It is very interesting to understand this behavior as either q and D vary. If D is fixed and q varies then obviously the number of points is fixed and the problem is really about the sup-norm at special points. As q and D vary together then we are studying a hybrid version. One can see some pleasant analogies between the methods used here and in [BKY], especially their Theorem 1.7 which is a geodesic restriction bound.

The restriction to prime level q is made to simplify some arguments. Probably with some extra work one can show that our results hold for all square-free integers $q > 1$ and odd fundamental discriminants $-D < -4$ such that every prime divisor of q splits in $\mathbb{Q}(\sqrt{-D})$. For simplicity we also do not keep track of the t_f -dependence in our estimates, but it is clear from the proof that the dependence is polynomial.

When q is fixed, Michel and Venkatesh [MV2] proved an asymptotic formula with a power savings in D for the first moment in Theorem 1.1 and gave applications to nonvanishing. An analogous result for central derivatives was obtained by Templier [Te]. Earlier, Duke, Friedlander, and Iwaniec [DFI] obtained an asymptotic formula for the second moment of class group L -functions which in Theorem 1.1 corresponds to replacing f by an Eisenstein series. The novelty in Theorem 1.1 is that both q and D are allowed to vary in a fairly wide range. Our proof is inspired in many ways by [MV2]; see in particular their Remark

2.1 where they set up the following spectral approach. We begin by using a formula of Waldspurger/Zhang [Wal] [Z1] for $L(f \times \Theta_\chi, 1/2)$ to relate the first moment to the sum of a fixed automorphic function F evaluated on Heegner points. We then spectrally decompose this sum and analyze each part of the spectral contribution separately. The main part of our analysis involves the estimation of the contribution of the Maass forms. Roughly, we must estimate an expression of the form

$$\sum_{|t_g| \ll (qD)^\varepsilon} \langle F, g \rangle_q W_{D,g}$$

where g runs over an orthonormal basis of Maass cusp forms for $\Gamma_0(q)$, $\langle \cdot, \cdot \rangle_q$ is the Petersson inner product on $L^2(Y_0(q))$, and $W_{D,g}$ is the ‘‘Weyl sum’’ of g evaluated on a Galois orbit of Heegner points of discriminant $-D$ on $X_0(q)$. Watson’s formula [Wats] relates $|\langle F, g \rangle_q|^2$ to the triple product L -function $L(f \times f \times g, 1/2) = L(\text{sym}^2 f \times g, 1/2)L(g, 1/2)$ while the Waldspurger/Zhang formula relates $|W_{D,g}|^2$ to $L(g, 1/2)L(g \times \chi_D, 1/2)$. Thus one is naturally led to estimating mean values over quite different families of L -functions from those occurring in Theorem 1.1. Various applications of Hölder’s inequality are possible here (see the end of Section 6 for further discussion), and in particular one is led to estimating the average of $L(g \times \chi_D, 1/2)$ which is a GL_2 family with smaller conductors and degree than the original L -functions. If q is very small compared to D then the best we can do is apply a hybrid subconvexity estimate of [BH]. However, if q is somewhat large then it is advantageous to use the following which is of independent interest.

Theorem 1.5. *Let $T > 0$, $1 \leq M \leq T$, and suppose χ_D is a primitive quadratic Dirichlet character of conductor $|D|$. Then for $SL_2(\mathbb{Z})$, we have*

$$\sum_{T \leq t_g \leq T+M} \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)} + \int_T^{T+M} \frac{|L(\frac{1}{2} + it, \chi_D)|^2}{|\zeta(1 + 2it)|^2} dt \ll (TM + \sqrt{|D|})(|D|MT)^\varepsilon. \quad (1.5)$$

Under the additional assumption that q is prime with $(D, q) = 1$, we have for any $M \geq 1$

$$\sum_{|t_g| \leq M} \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)} \ll (qM^2 + \sqrt{|D|})(|D|Mq)^\varepsilon, \quad (1.6)$$

where the sum is over Maass newforms for $\Gamma_0(q)$. In (1.5) and (1.6), the implied constant depends on $\varepsilon > 0$ only.

We emphasize that D can be any fundamental discriminant in Theorem 1.5 and q can be any odd prime coprime to D (in contrast to Theorem 1.1).

Estimates for $L(g \times \chi_D, 1/2)$ and $L(1/2 + it, \chi_D)$ have been studied in various aspects. For instance, with g fixed (alternately, t fixed), Conrey and Iwaniec [CI] have proved the Weyl-type subconvexity bound $L(g \times \chi_D, 1/2) + |L(1/2 + it, \chi_D)|^2 \ll_{g,t} |D|^{1/3+\varepsilon}$ by bounding the third moment on average over g of level dividing $|D|$. The nonnegativity of the central values [KaSa] [Bi] is crucial for obtaining this bound since otherwise one cannot drop all but one term. Recently Michel and Venkatesh [MV3] have obtained a general subconvexity result on GL_2 in particular valid for any range of q , T , and D . For Dirichlet L -functions, Heath-Brown [H-B1] [H-B2] showed $L(1/2 + it, \chi) \ll (q(1 + |t|))^{3/16+\varepsilon}$; see also [HW] [Watt]. Blomer and Harcos [BH] have obtained a general hybrid result for GL_2 with the use of an amplifier. Our approach to Theorem 1.5 has some elements in common with work of Munshi [Mu]. On dropping all but ‘‘one term’’, we obtain

Corollary 1.6. *Suppose g is a Hecke-Maass cusp form for $SL_2(\mathbb{Z})$. Then*

$$L(g \times \chi_D, \frac{1}{2}) + |L(\frac{1}{2} + iT, \chi_D)|^2 \ll (T + |D|^{1/2})(T|D|)^\varepsilon. \quad (1.7)$$

Similarly, if g is a Hecke-Maass cusp form for $\Gamma_0(q)$ with q prime and $(q, D) = 1$, we have

$$L(g \times \chi_D, \frac{1}{2}) \ll (q + |D|^{1/2})(q|D|)^\varepsilon, \quad (1.8)$$

The conductor of the L -functions in (1.7) is $(|D|T)^2$ so this is a subconvexity bound for $T^{1+\delta} \leq |D| \leq T^A$ for any fixed $\delta > 0$ and $A > 0$. For the Dirichlet L -function, this estimate improves on Heath-Brown's hybrid bound $L(1/2 + it, \chi_D) \ll (|D|t)^{3/16+\varepsilon}$ provided $t^{5/3} \leq |D| \leq t^3$. The bound (1.7) is strongest when $|D| \asymp T^2$ in which case it gives a Weyl-type bound. Similarly, (1.8) is subconvex for $q^{3/2+\delta} \leq |D| \leq q^A$ with fixed $\delta, A > 0$.

Based on our spectral approach to Theorem 1.1, it is desirable to estimate the L^4 -norm of an L^2 -normalized Maass form \tilde{f} of large level q and say bounded Laplace eigenvalue (see [BKY] for investigations into the weight aspect). Recently, Blomer [Bl] showed a best-possible (up to q^ε) estimate on average for such forms, and observed that $\|\tilde{f}\|_4^4 \ll q^{-1/3+\varepsilon}$ by interpolating the sup-norm bound of [HaTe] with the L^2 -normalization. As an easy byproduct of our work here, we record the following

Proposition 1.7. *Suppose \tilde{f} is a Hecke-Maass newform of prime level q and spectral parameter t_f which is L^2 -normalized according to the inner product (2.1). Then we have*

$$\|\tilde{f}\|_4^4 = \int_{Y_0(q)} |\tilde{f}(z)|^4 \frac{dx dy}{y^2} \ll_{t_f} q^{-1/2+\varepsilon}. \quad (1.9)$$

For comparison, the Lindelöf Hypothesis would give $O(q^{-1+\varepsilon})$ as the bound in (1.9).

The form of the bound in Theorem 1.5 has some elements in common with a result of Kohnen and Sengupta [KoSe], who showed

$$\sum_{g \in B_k} \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)} \ll_{D, \varepsilon} k^{1+\varepsilon}, \quad (1.10)$$

where the sum is over even weight k holomorphic Hecke eigenforms for $SL_2(\mathbb{Z})$ and $(-1)^k D > 0$. However, with some extra work (see Section 11 below), their bound can be improved to

$$\sum_{g \in B_k} \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)} \ll_\varepsilon (k + \sqrt{|D|})(k|D|)^\varepsilon, \quad (1.11)$$

which has the same form as in Theorem 1.5. Their proof bypasses some of the analytic techniques in this paper and instead relies on the Shimura correspondence and explicit calculations of the Fourier expansion of half-integral weight Poincaré series (alternatively, the Petersson formula for half-integral weight). This approach gives an elegant and direct proof but it is difficult to generalize it to the Maass form setting.

2. PRELIMINARIES ON MAASS FORMS

Let q be a positive integer which is either 1 or a prime number. Let $h_1, h_2 : \mathbb{H} \rightarrow \mathbb{C}$ be $\Gamma_0(q)$ -invariant functions on the complex upper half-plane \mathbb{H} , and define the Petersson inner product

$$\langle h_1, h_2 \rangle_q := \int_{Y_0(q)} h_1(z) \overline{h_2(z)} \frac{dx dy}{y^2}, \quad (2.1)$$

where $Y_0(q) := \Gamma_0(q) \backslash \mathbb{H}$ is the open modular curve. The spectrum of $L^2(Y_0(q))$ has an orthonormal basis consisting of the constant function, Maass forms, and Eisenstein series corresponding to the cusps $\mathfrak{a} = 0, \infty$ of $\Gamma_0(q)$. An orthonormal basis for the cuspidal spectrum of $L^2(Y_0(q))$ is given by

$$\mathcal{B} := \mathcal{B}_q \cup \mathcal{B}_1^{(q)} \cup \mathcal{B}_1^*,$$

where \mathcal{B}_q is an orthonormal basis of Hecke-Maass newforms of weight 0 for $\Gamma_0(q)$, $\mathcal{B}_1^{(q)}$ is a basis of Hecke-Maass cusp forms of weight 0 for $SL_2(\mathbb{Z})$ which is orthonormal with respect to the inner product (2.1), and

$$\mathcal{B}_1^* := \{g_q : g \in \mathcal{B}_1^{(q)}\},$$

where (see [ILS, Proposition 2.6])

$$g_q(z) := \left(1 - \frac{q\lambda_g^2(q)}{(q+1)^2}\right)^{-1/2} \left(g(qz) - \frac{q^{1/2}\lambda_g(q)}{q+1}g(z)\right), \quad (2.2)$$

and $\lambda_g(n)$ is the n -th Hecke eigenvalue of g . We have

$$\lambda_g(q) = \pm q^{-1/2} \quad \text{when } g \in \mathcal{B}_q. \quad (2.3)$$

It is also convenient to use the notation $\mathcal{B}_1^{(1)}$ to denote the same basis as $\mathcal{B}_1^{(q)}$ but rescaled to be orthonormal with respect to the inner product (2.1) with $q = 1$.

Let $t_g := \sqrt{\lambda_g - \frac{1}{4}}$ denote the spectral parameter of $g \in \mathcal{B}$ where λ_g is the Laplace eigenvalue of g . Then by Weyl's law we have

$$\#\{g \in \mathcal{B}_q : |t_g| \leq T\} \ll qT^2 \quad \text{and} \quad \#\{g \in \mathcal{B}_1^{(1)} : t_g \leq T\} \ll T^2. \quad (2.4)$$

3. A FORMULA OF WALDSPURGER/ZHANG

Let f be an arithmetically normalized Hecke-Maass newform of weight 0 for $\Gamma_0(q)$ with spectral parameter t_f , and \tilde{f} be the L^2 -normalized newform $\langle \tilde{f}, \tilde{f} \rangle_q = 1$ corresponding to f . Note that $\tilde{f} = a_{\tilde{f}}(1)f$, and if q is a prime (see [Bl, eq. (2.9)]),

$$\langle f, f \rangle_q = \frac{L(\text{sym}^2 f, 1)}{2 \cosh(\pi t_f)} \frac{q^2}{q-1}. \quad (3.1)$$

Define the completed Rankin-Selberg L -function

$$\Lambda(f \times \Theta_\chi, s) := L_\infty(f \times \Theta_\chi, s) L(f \times \Theta_\chi, s),$$

where

$$L_\infty(f \times \Theta_\chi, s) = 4(2\pi)^{-2s} \Gamma(s + it_f) \Gamma(s - it_f).$$

Let $\tau \in \Lambda_D(q)$ be a Heegner point on $X_0(q)$. Then one has the following central value formula due to Waldspurger/Zhang [Wal] [Z1] (though see p. 647 of [HaMi] for the explicit form)

$$\Lambda(f \times \Theta_\chi, \frac{1}{2}) = \frac{4\langle f, f \rangle_q}{\sqrt{D}} \left| \sum_{\sigma \in \text{CL}_K} \chi(\sigma) \tilde{f}(\tau^\sigma) \right|^2. \quad (3.2)$$

Note that for q a prime, we can use (3.1) combined with $\cosh(\pi t)\Gamma(\frac{1}{2} + it)\Gamma(\frac{1}{2} - it) = \pi$ to write (3.2) as

$$L(f \times \Theta_{\chi, \frac{1}{2}}) = \nu(q) \frac{L(\text{sym}^2 f, 1)}{\sqrt{D}} \left| \sum_{\sigma \in \text{CL}_K} \chi(\sigma) \tilde{f}(\tau^\sigma) \right|^2, \quad (3.3)$$

where $\nu(q) = q^2/(q-1)$. For $q=1$, the formula (3.3) holds with $\nu(1) = 2$.

4. PROOF OF THEOREM 1.1

Here we give an overview of the proof of Theorem 1.1. Using the orthogonality relations for the characters $\widehat{\text{CL}}_K$, we obtain from (3.3) the average formula

$$M_f(D) := \sum_{\chi \in \widehat{\text{CL}}_K} L(f \times \Theta_{\chi, \frac{1}{2}}) = \nu(q) L(\text{sym}^2 f, 1) \frac{h(-D)}{\sqrt{D}} \sum_{\sigma \in \text{CL}_K} |\tilde{f}(\tau^\sigma)|^2. \quad (4.1)$$

Up to a scaling factor, we view the right hand side as the L^2 norm of \tilde{f} restricted to the Galois orbit of Heegner points.

Note that (4.1) is invariant under the choice of Galois orbit (when q is prime there are two such orbits) which is consistent with the fact that $|\tilde{f}|^2$ is invariant under the Fricke involution, and so are the central value of the L -functions.

For notational convenience, set $F(z) = |\tilde{f}(z)|^2$. Then spectrally decomposing F in $L^2(Y_0(q))$ yields

$$F(z) = \frac{\langle F, 1 \rangle_q}{\text{vol}(Y_0(q))} + \sum_{g \in \mathcal{B}} \langle F, g \rangle_q g(z) + \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{\infty} \langle F, E_{\mathfrak{a}}(\cdot, \frac{1}{2} + it) \rangle_q E_{\mathfrak{a}}(z, \frac{1}{2} + it) dt, \quad (4.2)$$

where

$$E_{\mathfrak{a}}(z, s) = \sum_{\gamma \in \Gamma_{\mathfrak{a}} \backslash \Gamma_0(q)} \text{Im}(\sigma_{\mathfrak{a}}^{-1} \gamma z)^s$$

is the real-analytic Eisenstein series corresponding to the cusp \mathfrak{a} of $\Gamma_0(q)$ (here $\sigma_{\mathfrak{a}\infty} = \mathfrak{a}$). From (4.1) we obtain

$$M_f(D) = \nu(q) L(\text{sym}^2 f, 1) \frac{h(-D)}{\sqrt{D}} \left[\frac{h(-D)}{\text{vol}(Y_0(q))} \langle F, 1 \rangle_q + \sum_{g \in \mathcal{B}} \langle F, g \rangle_q W_{D,g} + \sum_{\mathfrak{a}} \int_{-\infty}^{\infty} \langle F, E_{\mathfrak{a}}(\cdot, \frac{1}{2} + it) \rangle_q W_{D,\mathfrak{a}}(t) \frac{dt}{4\pi} \right], \quad (4.3)$$

where the hyperbolic Weyl sums are defined by

$$W_{D,g} := \sum_{\sigma \in \text{CL}_K} g(\tau^\sigma), \quad W_{D,\mathfrak{a}}(t) := \sum_{\sigma \in \text{CL}_K} E_{\mathfrak{a}}(\tau^\sigma, \frac{1}{2} + it).$$

Using $\text{vol}(Y_0(q)) = \frac{\pi(q+1)}{3}$, and $\langle F, 1 \rangle_q = 1$, we find that the constant eigenfunction in (4.3) gives the main term appearing in (1.1). Proposition 5.3 bounds the continuous spectrum which leads to a contribution to $M_f(D)$ of size $O(D^{5/12}(qD)^\varepsilon)$. The contribution from $g \in \mathcal{B}_1^{(q)} \cup \mathcal{B}_1^*$ give a bound of the same size as for the Eisenstein series, by Lemma 6.1. By Lemma 6.2, the contribution to $M_f(D)$ from \mathcal{B}_q gives

$$\ll (qD)^\varepsilon \min(qD^{7/16}, q^{3/4}D^{1/4} + q^{1/4}D^{1/2}), \quad (4.4)$$

as claimed in Theorem 1.1.

5. PERIOD INTEGRAL FORMULAS

In this section we will evaluate the magnitude of the Weyl sums $W_{D,g}$, the inner products $\langle F, g \rangle_q$, and the analogous quantities with the continuous spectrum. First suppose $g \in \mathcal{B}_q$. Applying (3.3) with $\chi = \chi_0$ the trivial class group character, we have

$$|W_{D,g}|^2 = \frac{\sqrt{D}L(g \times \chi_D, \frac{1}{2})L(g, \frac{1}{2})}{\nu(q)L(\text{sym}^2g, 1)}. \quad (5.1)$$

It turns out that a similar formula holds for $g \in \mathcal{B}_1^{(q)}$ and $g_q \in \mathcal{B}_1^*$. Using the nonnegativity of $L(g, 1/2)$ and $L(g \times \chi_D, 1/2)$ we shall deduce the following

Lemma 5.1. *For $g \in \mathcal{B}$ we have*

$$W_{D,g} = \theta_{g,D} \frac{D^{1/4} L(g \times \chi_D, \frac{1}{2})^{1/2} L(g, \frac{1}{2})^{1/2}}{q^{1/2} L(\text{sym}^2g, 1)^{1/2}}, \quad (5.2)$$

where $\theta_{g,D}$ is some complex number satisfying $|\theta_{g,D}| \leq 10$.

Here we abused notation slightly; if $g_q \in \mathcal{B}_1^*$ then W_{D,g_q} is given by (5.2) with $g \in \mathcal{B}_1^{(q)}$ (from which g_q is constructed) appearing on the right hand side.

Proof. If $g \in \mathcal{B}_1^{(q)}$ with $\langle g, g \rangle_q = 1$, let \tilde{g} be the form associated to g in $\mathcal{B}_1^{(1)}$, so that $\langle \tilde{g}, \tilde{g} \rangle_1 = 1$. Then $g = c\tilde{g}$ where $c^2 = 1/(q+1)$. Following the proof of Harcos and Michel [HaMi, Theorem 6], we find that

$$|W_{D,g}|^2 = c^2 \left| \sum_{\sigma \in \text{CL}_K} \tilde{g}(\tau^\sigma) \right|^2,$$

where $\{\tau^\sigma : \sigma \in \text{CL}_K\}$ is the set of Heegner points of discriminant $-D$ on the level 1 modular curve $X_0(1)$. Then by (3.2) with $\chi = \chi_0$,

$$|W_{D,g}|^2 = \frac{\sqrt{D}}{2(q+1)} \frac{L(g \times \chi_D, \frac{1}{2})L(g, \frac{1}{2})}{L(\text{sym}^2g, 1)}.$$

If $g_q \in \mathcal{B}_1^*$, a similar argument using (2.2) and the trivial bound $\lambda_g(q) \ll q^{\frac{1}{2}}$ yields (5.2) in this case also. \square

By Watson's formula [Wats], for $g \in \mathcal{B}_q$, we have

$$|\langle F, g \rangle_q|^2 = \left| \int_{Y_0(q)} |\tilde{f}(z)|^2 g(z) \frac{dx dy}{y^2} \right|^2 = \frac{1}{4q^2} \frac{\Lambda(f \times f \times g, \frac{1}{2})}{\Lambda(\text{sym}^2f, 1)^2 \Lambda(\text{sym}^2g, 1)}, \quad (5.3)$$

where $\Lambda(f \times f \times g, s)$ is the completed triple product L -function (see e.g. [Wats, chapter 4]). A calculation with the archimedean place (see Section 4 of [BKY] for a convenient reference) gives

Lemma 5.2. *Suppose $g \in \mathcal{B}_q$. Then*

$$\langle F, g \rangle_q = \theta_{f,g} q^{-1} \frac{L(\text{sym}^2f \times g, \frac{1}{2})^{1/2} L(g, \frac{1}{2})^{1/2}}{L(\text{sym}^2f, 1) L(\text{sym}^2g, 1)^{1/2}}, \quad (5.4)$$

where $\theta_{f,g}$ is some complex constant which satisfies the bound

$$|\theta_{f,g}| \ll \exp(-\pi r(t_f, t_g)), \quad r(t_f, t_g) = \begin{cases} 0, & \text{if } |t_g| \leq 2|t_f|, \\ |t_g - 2t_f|, & \text{if } |t_g| > 2|t_f|. \end{cases} \quad (5.5)$$

In fact a more precise estimate is possible but we are not concerned with the t_f -behavior in this paper. Formula (38) of [N1] extends (5.4) to $g \in \mathcal{B}_1^{(q)}$ and using the Fricke involution we see that a similar formula holds with $g_q \in \mathcal{B}_1^*$.

Next we consider the Eisenstein series. By unfolding (see Section 5 of [Bl]) we have

$$\langle F, E_a(\cdot, s) \rangle_q = |a_{\bar{f}}(1)|^2 \frac{2L(f \times f, s) \Gamma_{\mathbb{R}}(s) \Gamma_{\mathbb{R}}(s + 2it_f) \Gamma_{\mathbb{R}}(s - 2it_f)}{\zeta(2s) 2^{s+2} \Gamma_{\mathbb{R}}(s+1)}. \quad (5.6)$$

Since F is invariant under the Fricke involution that switches the cusps 0 and ∞ , both cusps give the same inner product in (5.6) (see Section 13.2 of [I2]). By (3.1) we conclude

$$\langle F, E_a(\cdot, \frac{1}{2} + it) \rangle_q = \theta_{f,t} q^{-1} \frac{|\zeta(\frac{1}{2} + it)| |L(\text{sym}^2 f, \frac{1}{2} + it)|}{L(\text{sym}^2 f, 1) |\zeta(1 + 2it)|}, \quad (5.7)$$

where $\theta_{f,t}$ satisfies (5.5) with t_g replaced by t .

By (10.30) of [DFI] which generalizes to level q a formula of Gross and Zagier [GZ, p. 248],

$$\sum_{\mathfrak{a}} W_{D,\mathfrak{a}}(t) = \theta_{t,D} \frac{D^{1/4} |\zeta(\frac{1}{2} + it)| |L(\frac{1}{2} + it, \chi_D)|}{q^{1/2} |\zeta(1 + 2it)|}, \quad (5.8)$$

where $\theta_{t,D}$ is some complex number satisfying $|\theta_{t,D}| \leq 10$. To be precise, we should remark that (10.30) of [DFI] is in a different form than (5.8); there the authors had an individual cusp on the left hand side, and the sum was over both Galois orbits of Heegner points. To derive (5.8) we use the fact that the Fricke involution switches the two Galois orbits, and also switches the two Eisenstein series, and so (5.8) follows.

At this point it is easy to establish the following estimate.

Proposition 5.3. *We have*

$$\sum_{\mathfrak{a}} \int_{-\infty}^{\infty} \langle F, E_a(\cdot, \frac{1}{2} + it) \rangle_q W_{D,\mathfrak{a}}(t) \frac{dt}{4\pi} \ll_{t_f, \varepsilon} q^{-1+\varepsilon} D^{\frac{5}{12}+\varepsilon}. \quad (5.9)$$

Proof. By the rapid decay of $\theta_{f,t}$ with $|t| > (qD)^\varepsilon$, we may truncate the t -integral at this point with an acceptable error term. By (5.7), (5.8), the convexity bound $L(\text{sym}^2 f, 1/2 + it) \ll q^{1/2+\varepsilon}$, the Conrey and Iwaniec [CI] bound $L(1/2 + it, \chi_D) \ll D^{1/6}(qD)^\varepsilon$ (for t small), and standard [Ti] lower bounds on $|\zeta(1 + 2it)|$, we immediately obtain (5.9). \square

6. CONTRIBUTION OF THE SPECTRUM \mathcal{B}

In this section we analyze the cusp form sum in (4.3). The first step is to finitize the sum over g . Using self-adjointness of the hyperbolic Laplacian $\Delta = -y^2(\partial_x^2 + \partial_y^2)$, Stokes' theorem, the Cauchy-Schwarz inequality, and the calculations in Section 5 of [Bl], we have

$$\langle F, g \rangle_q = \left(\frac{1}{4} + t_g^2\right)^{-A} \langle \Delta^A F, g \rangle_q \leq \left(\frac{1}{4} + t_g^2\right)^{-A} \|\Delta^A F\|_2 \|g\|_2 \ll_{t_f, A} q^{1+\varepsilon} (1 + |t_g|)^{-A} \quad (6.1)$$

for each integer $A \geq 0$. Alternatively, one could apply Watson's formula [Wats] (for g a newform) or Nelson's extension [N1] for $g \in \mathcal{B}_1^{(q)} \cup \mathcal{B}_1^*$ to obtain this type of bound. Since

$\langle F, g \rangle_q$ and $W_{D,g}$ grow polynomially in q , D , etc. we may impose the truncation $|t_g| \leq (qD)^\varepsilon$ with a very small error term.

First we deal with the oldforms by showing

Lemma 6.1. *We have*

$$\sum_{\substack{g \in \mathcal{B}_1^{(q)} \cup \mathcal{B}_1^* \\ |t_g| \leq (qD)^\varepsilon}} \langle F, g \rangle_q W_{D,g} \ll q^{-1+\varepsilon} D^{\frac{5}{12}+\varepsilon}. \quad (6.2)$$

Proof. First we note that for $g \in \mathcal{B}_1^{(q)}$, we have $\sup_{z \in \mathbb{H}} |g(z)| \ll t_g^{1/2} q^{-1/2}$ where $t_g^{1/2}$ is the trivial bound on the sup norm of a level 1 Hecke cusp form, and the $q^{-1/2}$ comes from the normalization with respect to $Y_0(q)$. Similarly, inspection of (2.2) shows that the same bound holds for $g_q \in \mathcal{B}_1^*$. Thus for $g \in \mathcal{B}_1^{(q)} \cup \mathcal{B}_1^*$, we have $|\langle F, g \rangle_q| \ll t_g^{1/2} q^{-1/2} \langle F, 1 \rangle_q = t_g^{1/2} q^{-1/2}$. For the Weyl sum, we note by Lemma 5.1 and the Conrey and Iwaniec bound [CI], we have for $|t_g| \leq (qD)^\varepsilon$ that $W_{D,g} \ll q^{-1/2} D^{5/12} (qD)^\varepsilon$. Using the fact that there are $O((qD)^{2\varepsilon})$ oldforms with $t_g \leq (qD)^\varepsilon$, we finish the proof. \square

Note that the oldforms contribute the same amount as the Eisenstein series (compare Lemma 6.1 with Proposition 5.3). The main work remains to estimate the newforms.

Lemma 6.2. *We have*

$$\sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \langle F, g \rangle_q W_{D,g} \ll (qD)^\varepsilon \min(D^{7/16}, \frac{D^{1/4}}{q^{1/4}} + \frac{D^{1/2}}{q^{3/4}}). \quad (6.3)$$

Proof of Lemma 6.2. Combining (5.4) and (5.2) we obtain

$$\begin{aligned} \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \langle F, g \rangle_q W_{D,g} &= \frac{D^{1/4}}{q^{3/2} L(\text{sym}^2 f, 1)} \\ &\times \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{\theta'_{g,f,D}}{L(\text{sym}^2 g, 1)} L(g \times \chi_D, \frac{1}{2})^{\frac{1}{2}} L(g, \frac{1}{2}) L(\text{sym}^2 f \times g, \frac{1}{2})^{\frac{1}{2}}. \end{aligned} \quad (6.4)$$

We apply Hölder's inequality with exponents 2, 4, 4, respectively, obtaining

$$\sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \langle F, g \rangle_q W_{D,g} \ll \frac{D^{1/4}}{q^{3/2} L(\text{sym}^2 f, 1)} M_1^{1/2} M_2^{1/4} M_3^{1/4}, \quad (6.5)$$

where we have used nonnegativity of central values, and where

$$\begin{aligned} M_1 &= \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)}, & M_2 &= \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{L(g, \frac{1}{2})^4}{L(\text{sym}^2 g, 1)}, \\ M_3 &= \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{L(\text{sym}^2 f \times g, \frac{1}{2})^2}{L(\text{sym}^2 g, 1)}. \end{aligned} \quad (6.6)$$

For M_2 , it is a standard application of the spectral large sieve inequality that

$$M_2 \ll q(qD)^\varepsilon. \quad (6.7)$$

With similar technology combined with some deep inputs on the automorphy of Rankin-Selberg convolutions, we will show

Proposition 6.3. *We have*

$$M_3 \ll q^2(qD)^\varepsilon. \quad (6.8)$$

For M_1 , we have two different approaches. For q small compared to D we simply apply the best-known progress towards Lindelöf for $L(g \times \chi_D, 1/2)$ and multiply by the number of forms (i.e. $q(qD)^\varepsilon$). Currently the best result is [BH] which gives $L(g \times \chi_D, 1/2) \ll q^{1/2} D^{3/8} (qD)^\varepsilon$. For q larger we appeal to (1.6). Taken together, we obtain

$$M_1 \ll (qD)^\varepsilon \min(q^{3/2} D^{3/8}, q + D^{1/2}). \quad (6.9)$$

Taking these estimates for granted, we obtain (6.3) after a short calculation. \square

Now we discuss an alternate arrangements of Hölder's inequality which may be of interest. Applying Hölder's inequality in (6.4) with exponents 4, 4, 2, respectively, we obtain

$$\sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \langle F, g \rangle_q W_{D,g} \ll \frac{D^{1/4}}{q^{3/2} L(\text{sym}^2 f, 1)} M_1'^{1/4} M_2'^{1/4} M_3'^{1/2}, \quad (6.10)$$

where

$$M_1' = \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{L(g \times \chi_D, \frac{1}{2})^2}{L(\text{sym}^2 g, 1)}, \quad M_3' = \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq (qD)^\varepsilon}} \frac{L(\text{sym}^2 f \times g, \frac{1}{2})}{L(\text{sym}^2 g, 1)}.$$

The spectral large sieve inequality (see Theorem 7.24 of [IK] for example) easily shows $M_1' \ll (q + q^{1/2} D)(qD)^\varepsilon$ and it seems likely that improvements are possible here using current technology. One may hope to show $M_3' \ll q(qD)^\varepsilon$ as this is a family with roughly q elements with conductors of size approximately q^4 ; the weight aspect analog of this estimate was shown in [BKY]. Conditional on this bound on M_3' , one would obtain

$$M_f(D) \ll (q^{1/2} D^{1/4} + q^{3/8} D^{1/2})(qD)^\varepsilon,$$

which would imply a subconvexity bound for any range of q and D except when one of q or D is fixed.

7. SPARSE EQUIDISTRIBUTION

Here we develop a natural formulation of equidistribution of Heegner points of level q and discriminant D as q is allowed to vary with D , say restricted by $q \leq D^\eta$ for some fixed $\eta > 0$. The basic difficulty is that the spaces $Y_0(q)$ are varying. Here we briefly recall the usual definition of equidistribution of say Heegner points of level 1. Suppose that $U(z) : \Gamma_0(1) \backslash \mathbb{H} \rightarrow \mathbb{R}$ is a smooth, compactly supported function on $Y_0(1)$. Then equidistribution means

$$\lim_{D \rightarrow \infty} \frac{1}{h(-D)} \sum_{\sigma \in \text{CL}_K} U(\tau^\sigma) = \int_{Y_0(1)} U(z) \frac{3}{\pi} \frac{dx dy}{y^2}.$$

What we will do is construct a sequence (actually, many possible different sequences) of functions U_q invariant on $\Gamma_0(q)$ with “isometric” analytic properties, and measure equidistribution with such functions. To construct these functions, begin with a function $u : \mathbb{H} \rightarrow \mathbb{R}$ that is smooth with support on a compact set S such that no two points of S are $\Gamma_0(1)$ -equivalent (this condition is not strictly necessary). Then $U(z) = \sum_{\gamma \in \Gamma_0(1)} u(\gamma z)$ (a sum with at most one nonzero term) satisfies the above properties. Now for each q , choose $\omega_q \in \Gamma_0(1)/\Gamma_0(q)$ and define $U_q(z) = \sum_{\gamma \in \Gamma_0(q)} u(\omega_q \gamma z)$. In other words, we have $U_q(z) = U(z)$, if $z \in \Gamma_0(q)\omega_q^{-1}\Gamma_0(1)\backslash\mathbb{H}$, and $U_q(z) = 0$ otherwise. One can picture what is going on by taking $S \subset \mathcal{F}$ where \mathcal{F} is the usual fundamental domain for $\Gamma_0(1)\backslash\mathbb{H}$. Then for each q choose a fundamental domain \mathcal{F}_q for $\Gamma_0(q)\backslash\mathbb{H}$. Under the natural projection $\pi : \mathcal{F}_q \rightarrow \mathcal{F}$, the set S pulls back to $q + 1$ translates of S ; the construction of U_q is to choose one of these copies to be the support of U_q (restricted to \mathcal{F}_q), and U_q restricted to this copy is identical to U . This construction appears (to the authors) to be a natural way to maintain consistent choices of test function for varying q 's. For example, we have obvious statements like

$$\left(\int_{Y_0(1)} |U(z)|^\rho \frac{dx dy}{y^2} \right)^{1/\rho} = \left(\int_{Y_0(q)} |U_q(z)|^\rho \frac{dx dy}{y^2} \right)^{1/\rho},$$

for any $\rho > 0$ (including $\rho = \infty$). Furthermore, as $q \rightarrow \infty$, the measure of the support of U_q is shrinking compared to the total measure of \mathcal{F}_q so that we are capturing some notion of sparsity. Now we *define* joint equidistribution to mean

$$\frac{1}{h(-D)} \sum_{\sigma \in \text{CL}_K} U_q(\tau^\sigma) = \frac{1}{\text{vol}(Y_0(q))} \int_{Y_0(1)} U(z) \frac{dx dy}{y^2} + E(q, D), \quad (7.1)$$

where as $D \rightarrow \infty$ with $q = q(D) \leq D^n$, we require $E(q, D) = o(q^{-1})$. In practice we can only expect (7.1) to hold true for q small enough compared to D . In particular, by volume considerations we cannot even expect the Heegner points to be dense in $Y_0(q)$ unless $q = o(h(-D))$.

Theorem 7.1. *Let U and U_q be defined as above. Then*

$$\frac{1}{h(-D)} \sum_{\sigma \in \text{CL}_K} U_q(\tau^\sigma) = \frac{1}{\text{vol}(Y_0(q))} \int_{Y_0(1)} U(z) \frac{dx dy}{y^2} + O(q^{1/4} D^{-1/16+\varepsilon}), \quad (7.2)$$

and therefore the Heegner points of level q become equidistributed in $Y_0(q)$ provided $q \leq D^{1/20-\varepsilon}$. The implied constant depends on U and ε but not on the choice of ω_q 's.

Proof. The proof follows similar lines to Theorem 1.1. We begin with the spectral decomposition

$$U_q(z) = \frac{\langle U_q, 1 \rangle_q}{\text{vol}(Y_0(q))} + \sum_{g \in \mathcal{B}} \langle U_q, g \rangle_q g(z) + \frac{1}{4\pi} \sum_a \int_{-\infty}^{\infty} \langle U_q, E_a(\cdot, \frac{1}{2} + it) \rangle_q E_a(z, \frac{1}{2} + it) dt. \quad (7.3)$$

Note $\langle U_q, 1 \rangle_q = \langle U, 1 \rangle_1$. Then inserting (7.3) into (7.2), we obtain

$$\sum_{\sigma \in \text{CL}_K} U_q(\tau^\sigma) = \frac{h(-D) \langle U, 1 \rangle_1}{\text{vol}(Y_0(q))} + \sum_{g \in \mathcal{B}} \langle U_q, g \rangle_q W_{D,g} + \sum_a \int_{-\infty}^{\infty} \langle U_q, E_a(\cdot, \frac{1}{2} + it) \rangle_q W_{D,a}(t) \frac{dt}{4\pi}. \quad (7.4)$$

The analog of (6.1) shows that we can truncate the spectral sum (and integral) at $|t_g|, |t| \leq D^\varepsilon$, with a very small error term. Then we apply Cauchy-Schwarz and Bessel's inequality to

obtain

$$\left| \sum_{\sigma \in \text{CL}_K} U_q(\tau^\sigma) - \frac{h(-D)\langle U, 1 \rangle_1}{\text{vol}(Y_0(q))} \right|^2 \leq \langle U_q, U_q \rangle_q \left(\sum_{\substack{g \in \mathcal{B} \\ |t_g| \leq D^\varepsilon}} |W_{D,g}|^2 + \sum_{\mathfrak{a}} \int_{-D^\varepsilon}^{D^\varepsilon} |W_{D,\mathfrak{a}}(t)|^2 dt \right) + O(D^{-100}). \quad (7.5)$$

Then by Lemma 5.1, (5.8), the [BH] bound of $L(g \times \chi_D, \frac{1}{2}) \ll q^{1/2} D^{3/8} (qD)^\varepsilon$, and the first moment bound (compare to (6.7))

$$\sum_{\substack{g \in \mathcal{B} \\ |t_g| \leq D^\varepsilon}} \frac{L(g, 1/2)}{L(\text{sym}^2 g, 1)} \ll q D^\varepsilon,$$

as well as easier analogues for the continuous spectrum, we have

$$\left| \sum_{\sigma \in \text{CL}_K} U_q(\tau^\sigma) - \frac{h(-D)\langle U, 1 \rangle_1}{\text{vol}(Y_0(q))} \right|^2 \ll_U q^{1/2} D^{7/8+\varepsilon}. \quad (7.6)$$

Then with some simplifications we complete the proof. \square

8. PROOF OF PROPOSITION 6.3

In this section we prove Proposition 6.3. The basic idea is to apply the spectral large sieve inequality. We begin by collecting some standard facts, starting with the spectral large sieve inequality.

Proposition 8.1. *Let $\{u_j\}$ be an orthonormal basis of Maass cusp forms for $\Gamma_0(q)$. Let $\lambda_j(n)$ be the n -th Hecke eigenvalue of u_j . Let $T \geq 1$ and $N \geq 1$. Then for any complex numbers $\{a_n\}_{n=1}^N$, we have*

$$\sum_{t_j \leq T} \frac{1}{L(\text{sym}^2 u_j, 1)} \left| \sum_{n=1}^N a_n \lambda_j(n) \right|^2 \ll (qT^2 + N \log(N)) (qT)^\varepsilon \sum_{n \leq N} |a_n|^2. \quad (8.1)$$

By Gelbart and Jacquet [GJ] the symmetric square lift $\text{sym}^2 f$ is a self-dual automorphic form on GL_3 with Fourier coefficients $A(m, k)$ satisfying

$$A(m, 1) = \sum_{ab^2=m} \lambda_f(a^2) \quad \text{when } q \nmid m,$$

and

$$A(m, k) = \sum_{d|(m,k)} \mu(d) A\left(\frac{m}{d}, 1\right) A\left(1, \frac{k}{d}\right) \quad \text{when } q \nmid mk.$$

Xiannan Li [L] showed the following uniform bound

$$\sum_{mk^2 \leq N} \frac{|A(m, k)|^2}{mk} \ll_{t_f} (qN)^\varepsilon. \quad (8.2)$$

We have

$$\begin{aligned} L(\text{sym}^2 f \times g, s) &= (1 - \lambda_g(q)q^{-s})^{-1}(1 - \lambda_g(q)q^{-(s+1)})^{-1}L^{(q)}(\text{sym}^2 f \times g, s) \\ &=: \sum_{n=1}^{\infty} \frac{\lambda_{\text{sym}^2 f \times g}(n)}{n^s}, \end{aligned} \quad (8.3)$$

where

$$L^{(q)}(\text{sym}^2 f \times g, s) = \sum_{(mk, q)=1} \frac{A(m, k)\lambda_g(m)}{(mk^2)^s}.$$

The conductor of $L(\text{sym}^2 f \times g, 1/2)$ is q^4 which requires a brief explanation. Watson calculated that the conductor of $L(f \times f \times g, 1/2)$ is q^5 (see p. 45 of [Wats]), while the conductor of $L(g, 1/2)$ is q , so using the factorization $L(f \times f \times g, s) = L(\text{sym}^2 f \times g, s)L(g, s)$, we make this deduction. Using Lemma 3.4 of [LY] (which is a useful variant on the approximate functional equation), we have

$$M_3 \ll q^\varepsilon \int_{-\log q}^{\log q} \sum_{\substack{|t_g| \leq (qD)^\varepsilon \\ L(\text{sym}^2 g, 1)}} \frac{1}{L(\text{sym}^2 g, 1)} \left| \sum_{n \ll q^{2+\varepsilon}} \frac{\lambda_{\text{sym}^2 f \times g}(n)}{n^{1/2+it}} W(n) \right|^2 + O((qD)^{-100}), \quad (8.4)$$

where $W(n)$ is some bounded function depending on q but not on t_g . In fact, (8.3) shows that it suffices to bound the sum over n coprime to q ; one way to see this is to follow the proof of Lemma 3.4 of [LY] and in their (3.34) factor the L -function as in (8.3) and trivially bound the contribution from the prime q .

By unraveling the definition of Dirichlet series coefficients, and using Cauchy's inequality, we obtain

$$\begin{aligned} \left| \sum_{\substack{(n, q)=1 \\ n \ll q^{2+\varepsilon}}} \frac{\lambda_{\text{sym}^2 f \times g}(n)}{n^{1/2+it}} W(n) \right|^2 &= \left| \sum_{\substack{(mk, q)=1 \\ mk^2 \ll q^{2+\varepsilon}}} \frac{A(m, k)\lambda_g(m)}{m^{1/2+it}k^{1+2it}} W(mk^2) \right|^2 \\ &\ll q^\varepsilon \sum_{k \ll q^{1+\varepsilon}} k^{-1} \left| \sum_{\substack{(m, q)=1 \\ m \ll q^{2+\varepsilon}/k^2}} \frac{A(m, k)\lambda_g(m)}{m^{1/2+it}} W(mk^2) \right|^2. \end{aligned} \quad (8.5)$$

Inserting (8.5) into (8.4) (after freely imposing the condition $(n, q) = 1$), and using Proposition 8.1, we obtain

$$M_3 \ll q^\varepsilon \sum_{k \ll q^{1+\varepsilon}} k^{-1} \left(q + \frac{q^2}{k^2} \right) \sum_{m \leq q^{2+\varepsilon}/k^2} \frac{|A(m, k)|^2}{m}. \quad (8.6)$$

Then using (8.2) completes the proof of Proposition 6.3.

9. PROOF OF PROPOSITION 1.7

Next we give the proof of Proposition 1.7 which with our current notation gives an upper bound on $\langle F, F \rangle_q$. By Parseval, we have

$$\langle F, F \rangle_q = \sum_{g \in \mathcal{B}} |\langle F, g \rangle_q|^2 + \dots,$$

where the dots indicate the continuous spectrum as well as the constant eigenfunction (which is easily seen to give $O(q^{-1+\varepsilon})$). As in the proof of Lemma 6.1, the sup-norm of an oldform is $O(q^{-1/2})$ so that these terms also give $O(q^{-1+\varepsilon})$. By (5.4) and Cauchy's inequality, we have

$$\sum_{g \in \mathcal{B}_q} |\langle F, g \rangle_q|^2 \ll q^{-2} M_3^{1/2} M_4^{1/2}, \quad M_4 = \sum_{\substack{g \in \mathcal{B}_q \\ |t_g| \leq q^\varepsilon}} \frac{L(g, 1/2)^2}{L(\text{sym}^2 g, 1)}, \quad (9.1)$$

where M_3 is as in (6.6). Then by Proposition 6.3 and the bound $M_4 \ll q^{1+\varepsilon}$ (implied by (6.7)), we obtain the bound of $O(q^{-1/2+\varepsilon})$ for the newforms. By (5.7), and the convexity bound $L(\text{sym}^2 f, 1/2 + it) \ll q^{1/2+\varepsilon}$ (for $t \ll q^\varepsilon$), we see that the Eisenstein series contribute $O(q^{-1+\varepsilon})$ just like the oldforms. This completes the proof.

10. PROOF OF THEOREM 1.5

We shall treat both bounds in Theorem 1.5 simultaneously as much as possible. For (1.6) we cover the set of $|t_g| \leq M$ by the set $|t_g| \leq 1$ and $O(M^\varepsilon)$ subintervals of the form $T \leq t_g \leq T + M'$ with $1 \leq T \leq M$, and $M' = T^{1-\varepsilon}$.

In the $SL_2(\mathbb{Z})$ case we may assume $T \geq |D|^\varepsilon$ since otherwise the convexity bound applied to every term immediately gives (1.5). Let $h(t)$ be a smooth, even, nonnegative function satisfying

$$\begin{cases} h(t) \text{ extends to a holomorphic function on } |\text{Im}(t)| \leq A \\ h(t) \ll (1 + |t|)^{-100}, \\ h(\pm \frac{(2n+1)i}{2}) = 0 \text{ for } 0 \leq n \leq A, \end{cases} \quad (10.1)$$

where A is some large positive parameter depending on $\varepsilon > 0$ desired in Theorem 1.5. We shall eventually take

$$h(t) = P(t) \left[\exp\left(-\left(\frac{t-T}{M}\right)^2\right) + \exp\left(-\left(\frac{t+T}{M}\right)^2\right) \right], \quad (10.2)$$

where $P(t)$ is an even polynomial vanishing at $i/2, 3i/2, \dots$. We also use (10.2) for $T = 0$, $M = 1$ to handle $|t| \leq 1$. For instance, we may take

$$P(t) = c \frac{(t^2 + \frac{1}{4})}{T^2 + \frac{1}{4}} \frac{(t^2 + \frac{3}{4})}{T^2 + \frac{3}{4}} \dots, \quad (10.3)$$

where c is a constant independent of t . We furthermore suppose $h(t) \geq 1$ for t in the region of interest, i.e., $T \leq t \leq T + M$ or $-\frac{1}{4} \leq it \leq \frac{1}{4}$. Then in (1.5) or (1.6) we can attach the smooth weight h and extend the sum (or integral) to all t_g (or t), for purposes of obtaining an upper bound. Then define for $SL_2(\mathbb{Z})$,

$$\mathcal{M}_1 = \sum_{g \in \mathcal{B}_1^{(1)}} h(t_g) \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)} + \frac{1}{\pi} \int_{-\infty}^{\infty} h(t) \frac{|L(\frac{1}{2} + it, \chi_D)|^2}{|\zeta(1 + 2it)|^2} dt, \quad (10.4)$$

and for the level q case,

$$\mathcal{M}_q = \sum_{g \in \mathcal{B}_q} h(t_g) \frac{L(g \times \chi_D, \frac{1}{2})}{L(\text{sym}^2 g, 1)}. \quad (10.5)$$

Our next step is to use a ‘‘long’’ (relative to the size of the conductor) one-piece approximate functional equation for the L -functions under consideration.

Proposition 10.1. *Fix an integer $d \geq 1$. There exists a smooth function $V(x)$ such that for any L -function $L(f, s) = \sum_{n=1}^{\infty} \lambda_f(n)n^{-s}$ (as in Chapter 5 [IK]) of degree d and with analytic conductor $\leq Q$, we have*

$$L(f, \frac{1}{2}) = \sum_{n=1}^{\infty} \frac{\lambda_f(n)}{\sqrt{n}} V(n/X) + O_{A,d}((X/Q)^{-A}), \quad (10.6)$$

where $A > 0$ is arbitrary and the implied constant depends only on A and d .

Proof. (Sketch) Let

$$V(x) = \frac{1}{2\pi i} \int_{(1)} \frac{\Gamma(2d(s+A+1))}{\Gamma(2d(A+1))} x^{-s} \frac{ds}{s}. \quad (10.7)$$

Then

$$\sum_{n=1}^{\infty} \frac{\lambda_f(n)}{\sqrt{n}} V(n/X) = \frac{1}{2\pi i} \int_{(1)} X^s \frac{\Gamma(2d(s+A+1))}{\Gamma(2d(A+1))} L(f, 1/2+s) \frac{ds}{s}. \quad (10.8)$$

Shifting contours to $\text{Re}(s) = -A$ picks up the value $L(f, 1/2)$ from the pole at $s = 0$, and using the functional equation and (5.114) of [IK], we obtain the desired estimate by a trivial bound on the new contour of integration. \square

The conductor of $L(g \times \chi_D, 1/2)$ is $qD^2(1 + |t_g|^2)$, so we set

$$Q = qD^2(T+1)^2, \quad \text{and} \quad X = Q^{1+\varepsilon}. \quad (10.9)$$

Then we have

$$\mathcal{M}_1 = \sum_{n=1}^{\infty} \frac{\chi_D(n)}{\sqrt{n}} V(n/X) \left(\sum_{g \in \mathcal{B}_1^{(1)}} \frac{h(t_g)}{L(\text{sym}^2 g, 1)} \lambda_g(n) + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{h(t)}{|\zeta(1+2it)|^2} \tau_{it}(n) dt \right) + O(Q^{-100}),$$

and similarly for \mathcal{M}_q . It is very convenient that our method allows us to use this one-piece approximate functional equation because then we do not need to split the family into pieces depending on the parity of g , χ_D , etc.

Recall that for $g \in \mathcal{B}_q \cup \mathcal{B}_1^{(q)}$,

$$\frac{1}{L(\text{sym}^2 g, 1)} \asymp q \frac{|\rho_g(1)|^2}{\cosh(\pi t_g)},$$

where $\rho_g(n)$ is the n -th Fourier coefficient of g when it is L^2 -normalized on $Y_0(q)$; see (2.9) of [Bl]. Thus we have for $l = 1$ or q , with the notation $\mathcal{B}_q^{(q)}$ denoting \mathcal{B}_q ,

$$\mathcal{M}_l \ll l \sum_{n=1}^{\infty} \frac{\chi_D(n)}{\sqrt{n}} V(n/X) \left(\sum_{g \in \mathcal{B}_l^{(l)}} \frac{h(t_g)}{\cosh(\pi t_g)} \rho_g(n) \overline{\rho_g(1)} + (\text{continuous}) \right) + O(Q^{-100}). \quad (10.10)$$

The continuous spectrum contribution to \mathcal{M}_q is nonnegative. Next, for the case $l = q$, we wish to extend the spectral sum to a full orthonormal basis for $L^2(Y_0(q))$, not just newforms; that is, we include $\mathcal{B}_1^{(q)}$ and \mathcal{B}_1^* . We claim that this inclusion only increases the right hand side of (10.10), up to a negligible error term. It is natural to combine the forms $g \in \mathcal{B}_1^{(q)}$ and the corresponding form $g_q \in \mathcal{B}_1^*$. Recall from (2.2) that $g_q(z) = ag(z) + bg(qz)$ for $g \in \mathcal{B}_1^{(q)}$,

where $|a| \asymp \frac{|\lambda_g(q)|}{\sqrt{q}}$, and $|b| \asymp 1$. Thus $\rho_{g_q}(n) = a\rho_g(n) + b\rho_g(n/q)$ (the latter term denoting zero if $q \nmid n$), so that

$$\sum_{n=1}^{\infty} \frac{\rho_g(n)\overline{\rho_g(1)}\chi_D(n)}{\sqrt{n}}V(n/X) + \sum_{n=1}^{\infty} \frac{\rho_{g_q}(n)\overline{\rho_{g_q}(1)}\chi_D(n)}{\sqrt{n}}V(n/X)$$

simplifies as

$$|\rho_g(1)|^2(1+a^2)\sum_{n=1}^{\infty} \frac{\lambda_g(n)\chi_D(n)}{\sqrt{n}}V(n/X) + |\rho_g(1)|^2 \frac{ab\chi_D(q)}{\sqrt{q}} \sum_{n=1}^{\infty} \frac{\lambda_g(n)\chi_D(n)}{\sqrt{n}}V\left(\frac{n}{X/q}\right).$$

Using (10.6) again (in reverse), we have that this becomes

$$|\rho_g(1)|^2 \left(1 + a^2 + \frac{ab\chi_D(q)}{\sqrt{q}}\right) L(g \times \chi_D, 1/2) + O(Q^{-100}),$$

where we use the fact that X/q is still larger than the conductor of $L(g \times \chi_D, 1/2)$ since g is level 1. Standard bounds on $\lambda_g(q)$ show that $(1 + a^2 \pm \frac{ab}{\sqrt{q}}) \gg 1$ with an absolute implied constant, so by positivity of the central values again we see that the claim is proved.

Next we require the Kuznetsov formula.

Lemma 10.2. *We have*

$$\begin{aligned} \sum_g \overline{\rho_g(m)}\rho_g(n) \frac{h(t_g)}{\cosh(\pi t_g)} + \sum_a \frac{1}{4\pi} \int_{-\infty}^{\infty} \tau_a(m, t)\tau_a(n, t) \frac{h(t)}{\cosh(\pi t)} dt \\ = \delta_{m,n}H_0 + \sum_{c \equiv 0 \pmod{q}} \frac{S(m, n; c)}{c} H\left(\frac{4\pi\sqrt{mn}}{c}\right), \end{aligned} \quad (10.11)$$

where the sum over g runs over an orthonormal basis of Maass forms for $\Gamma_0(q)$,

$$H_0 = \pi^{-2} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr \quad (10.12)$$

and

$$H(y) = \frac{2i}{\pi} \int_{-\infty}^{\infty} J_{2ir}(y) \frac{r h(r)}{\cosh(\pi r)} dr. \quad (10.13)$$

We then have for $l = 1, q$,

$$M_l \ll lH_0 + lS_l + O(Q^{-100}), \quad S_l = \sum_{n=1}^{\infty} \frac{\chi_D(n)}{\sqrt{n}}V(n/X) \sum_{c \equiv 0 \pmod{q}} \frac{S(n, 1; c)}{c} H\left(\frac{4\pi\sqrt{n}}{c}\right). \quad (10.14)$$

An easy calculation shows $H_0 \ll (T+1)M$, consistent with Theorem 1.5 (after adding up $\ll X^\epsilon$ such intervals).

The proof of Theorem 1.5 then reduces to showing the following

Proposition 10.3. *We have*

$$S_l \ll l^{-1} \sqrt{|D|} X^\epsilon. \quad (10.15)$$

The proof of Proposition 10.3 requires some auxiliary lemmas. We presently develop some properties of $H(y)$.

Lemma 10.4. *Suppose h satisfies (10.1). Then*

$$H^{(j)}(y) \ll_{j,A} (T+1)M \left(\frac{y}{T+1} \right)^A, \quad j = 0, 1, 2, \dots, 200. \quad (10.16)$$

Furthermore, if h is of the form (10.2), there exist functions u_{\pm} satisfying the derivative bound

$$\frac{d^j}{dv^j} u_{\pm}(v) \ll_{j,A} (1+|v|)^{-A}, \quad (10.17)$$

so that

$$H(y) = (T+1) \sum_{\delta=\pm} \int_{-\infty}^{\infty} e(\delta T v/M) \cos(y \cosh(\pi v/M)) u_{\delta}(v) dv. \quad (10.18)$$

Proof. For (10.16), we move the contour of integration in (10.13) so $\operatorname{Re}(2ir) = A$ and use the bound

$$J_{A+2iv}(y) \ll_A \left(\frac{y}{1+|v|} \right)^A \exp(\pi|v|), \quad (10.19)$$

which follows from the integral representation ([GR] 8.411.4)

$$J_{\nu}(y) = 2 \frac{(y/2)^{\nu}}{\Gamma(\nu + \frac{1}{2})\Gamma(\frac{1}{2})} \int_0^{\pi/2} (\sin \theta)^{2\nu} \cos(y \cos \theta) d\theta, \quad \operatorname{Re}(\nu) > -\frac{1}{2}.$$

The estimates for the derivatives of H follow similarly by the use of the formula

$$\frac{d}{dz} J_{\nu}(z) = \frac{1}{2} J_{\nu-1}(z) - \frac{1}{2} J_{\nu+1}(z)$$

and changing variables $r \rightarrow r \pm i/2$. This process shows that $H^{(j)}$ is given by an integral representation similar to that of H but with a slightly different kernel function $rh(r)/\cosh(\pi r)$ replaced by linear combinations of $(r \pm ki/2)h(r \pm ki/2)/\cosh(\pi(r \pm ki/2))$ with $k \in \{-200, -199, \dots, 200\}$. Thus (10.16) holds for $1 \leq j \leq 200$ also.

Next we show (10.18). Using the fact that h is even and 8.411.11 of [GR], that is

$$J_{\nu}(y) = \frac{2}{\pi} \int_0^{\infty} \sin \left(y \cosh t - \frac{\nu\pi}{2} \right) \cosh(\nu t) dt,$$

we derive the integral representation

$$H(y) = \frac{2}{\pi^2} \int_{-\infty}^{\infty} r \tanh(\pi r) h(r) \int_{-\infty}^{\infty} e\left(\frac{rv}{\pi}\right) \cos(y \cosh v) dv dr. \quad (10.20)$$

Integrating by parts once in the inner v -integral shows that for $V \geq 1$

$$\int_{|v| \geq V} e\left(\frac{rv}{\pi}\right) \cos(y \cosh v) dv \ll \frac{1+|r|}{y \exp(V)}.$$

Then by interchanging the orders of integration, we have

$$H(y) = \int_{|v| \leq V} \cos(y \cosh v) U(v) dv + O((T+1)^2 M y^{-1} \exp(-V)), \quad (10.21)$$

where

$$U(v) = \int_{-\infty}^{\infty} \frac{2}{\pi^2} r \tanh(\pi r) h(r) e\left(\frac{rv}{\pi}\right) dr. \quad (10.22)$$

Using the formula (10.2), and changing variables, we have

$$U(v) = M(T+1)e\left(\frac{\delta T v}{\pi}\right) \frac{1}{\pi^2} \sum_{\delta=\pm} \int_{-\infty}^{\infty} \frac{Mr + \delta T}{T+1} P(Mr + \delta T) \tanh(\pi(Mr + \delta T)) e^{-r^2} e\left(\frac{Mvr}{\pi}\right) dr.$$

The integral becomes the Fourier transform, evaluated at $-Mv/\pi$, of a function $W(r) = w_{M,T,\delta}(r)$ satisfying

$$W^{(j)}(r) \ll_{j,A} (1 + |r|)^{-A} \quad (10.23)$$

and hence its Fourier transform \widehat{W} also satisfies (10.23). Changing variables $v \rightarrow v\pi/M$ and taking the limit as $V \rightarrow \infty$ then gives (10.18). \square

Now we begin the analysis of S_l . By applying a smooth dyadic partition of unity to the n -sum, we may assume $N \leq n \leq 2N$, where $1 \leq N \ll X$. Using (10.16) allows us to assume that $c \leq C$, where

$$C = \frac{\sqrt{N}}{T+1} X^\varepsilon. \quad (10.24)$$

Thus it suffices to show $S(N) \ll l^{-1} \sqrt{|D|} X^\varepsilon$, where

$$S(N) = \sum_{\substack{c \equiv 0 \pmod{q} \\ c \leq C}} c^{-1} N^{-\frac{1}{2}} S_c(N), \quad S_c(N) := \sum_{n=1}^{\infty} \chi_D(n) S(n, 1; c) w(n) H\left(\frac{4\pi\sqrt{n}}{c}\right), \quad (10.25)$$

where w has support in $[N, 2N]$ and satisfies $w^{(j)}(x) \ll N^{-j}$.

By Poisson summation in n modulo $c|D|$, we have

$$S_c(N) = \sum_{m \in \mathbb{Z}} \frac{1}{c|D|} a(m; c, D) r(m; c, D), \quad (10.26)$$

where

$$a(m; c, D) = \sum_{x \pmod{cD}} \chi_D(x) S(x, 1; c) e\left(\frac{mx}{c|D|}\right), \quad (10.27)$$

and

$$r(m; c, D) = \int_0^\infty w(x) H\left(\frac{4\pi\sqrt{x}}{c}\right) e\left(\frac{-mx}{c|D|}\right) dx. \quad (10.28)$$

Lemma 10.5. *Suppose that $c = c_1 c_2$ and $D = D_1 D_2$ with $(c_1 D_1, c_2 D_2) = 1$. Then*

$$a(m; c_1 c_2; D_1 D_2) = a(m c_2 \overline{D_2}; c_1, D_1) a(m c_1 \overline{D_1}; c_2, D_2). \quad (10.29)$$

Furthermore, we have the bound

$$|a(m; c, D)| \leq 4^{\nu+2} c |D|^{1/2} (m, c, D)^{1/2}, \quad (10.30)$$

where $c = 2^\nu c'$ with $(c', 2) = 1$. Finally, $a(0; c, D) = 0$ if $q|c$.

Lemma 10.6. *We have*

$$r(m; c, D) \ll N X^\varepsilon \left(1 + \frac{|m| \sqrt{N}}{|D|}\right)^{-2}. \quad (10.31)$$

Furthermore, if $T \geq X^\varepsilon$ then $r(0; c, D) = O(X^{-100})$.

We postpone the proof of these two lemmas and finish the proof of Proposition 10.3. In the $\Gamma_0(q)$ case (where the term $m = 0$ vanishes), we have

$$S(N) \ll X^\varepsilon \sum_{\substack{c \equiv 0 \pmod{q} \\ c \leq C}} \sum_{m \neq 0} \frac{1}{c\sqrt{N}} |D|^{-1/2}(c, D, m) N \left(\frac{m\sqrt{N}}{|D|} \right)^{-2} \ll X^\varepsilon \sqrt{|D|}/q. \quad (10.32)$$

The case of $\Gamma_0(1)$ gives an identical bound to (10.32) (with $q = 1$) since in this case the $m = 0$ term is practically negligible by Lemma 10.6.

Proof of Lemma 10.5. By the Chinese remainder theorem, write

$$x = x_1 c_2 D_2 \overline{c_2 D_2} + x_2 c_1 D_1 \overline{c_1 D_1},$$

where x_1 runs modulo $c_1 D_1$, x_2 runs modulo $c_2 D_2$, and $c_i \overline{c_i} \equiv D_i \overline{D_i} \equiv 1 \pmod{c_j D_j}$ with $i \neq j$. Then using $\chi_{D_1 D_2}(x) = \chi_{D_1}(x) \chi_{D_2}(x)$, we have

$$a(m; c, D) = \left(\sum_{x_1 \pmod{c_1 D_1}} \chi_{D_1}(x_1) S(x_1 \overline{c_2^2}, 1; c_1) e\left(\frac{m x_1 \overline{c_2 D_2}}{c_1 D_1}\right) \right) \times (\text{similar}),$$

where the term ‘‘similar’’ is identical to the first term but with c_1 switched with c_2 , D_1 switched with D_2 , and x_1 switched with x_2 . Changing variables $x_1 \rightarrow c_2^2 x_1$ and $x_2 \rightarrow c_1^2 x_2$, we immediately obtain (10.29).

We shall bound $a(m; c, D)$ for c and D powers of the same prime p ; this suffices by (10.29). If $p = 2$ we only claim the trivial bound for simplicity. Now suppose p is odd. By opening the Kloosterman sum, we have

$$a(m; p^c, \pm p^D) = \sum_{x \pmod{p^{c+D}}} \sum_{y \pmod{p^c}}^* \chi_{\pm p^D}(x) e\left(\frac{xy + \overline{y}}{p^c}\right) e\left(\frac{mx}{p^{c+D}}\right), \quad (10.33)$$

where $\pm p^D$ is a fundamental discriminant, or 1. We consider a variety of cases. Suppose $D = 0$ and $c \geq 1$. Then the sum over x vanishes unless $y \equiv -m \pmod{p^c}$ in which case necessarily $(m, p) = 1$. Hence

$$a(m; p^c, 1) = p^c e\left(\frac{-\overline{m}}{p^c}\right). \quad (10.34)$$

If $D = 1$ and $c = 0$ then we obtain a Gauss sum, so

$$a(m; 1, \pm p) = \epsilon_p \left(\frac{m}{p}\right) \sqrt{p}. \quad (10.35)$$

Suppose now that $c \geq 1$. Changing variables $x \rightarrow x + p^c$ shows that $a(m; p^c, \pm p) = e\left(\frac{m}{p}\right) a(m; p^c, \pm p)$ so it vanishes unless $p|m$. Accordingly, write $m = pm_1$. The sum over x is then periodic modulo p^c and is therefore the same sum repeated p times, so we have

$$a(m; p^c, \pm p) = p \sum_{x \pmod{p^c}} \sum_{y \pmod{p^c}}^* \left(\frac{x}{p}\right) e\left(\frac{xy + \overline{y}}{p^c}\right) e\left(\frac{m_1 x}{p^c}\right). \quad (10.36)$$

Suppose that $c = 1$. Then the sum over x is a Gauss sum, and we have

$$a(m; p, \pm p) = \epsilon_p p^{3/2} \sum_{y \pmod{p}}^* \left(\frac{m_1 + y}{p}\right) e\left(\frac{\overline{y}}{p}\right). \quad (10.37)$$

By [S] Theorem 2.6, the inner sum over y is bounded in absolute value by \sqrt{p} , so $|a(m; p, \pm p)| \leq p^2 = p^c p^{D/2} (m, p^c, p^D)^{1/2}$, as desired. For the purpose of proving Proposition 10.3 we could get away with using only the trivial bound in place of the Riemann Hypothesis for curves. Now suppose that $c \geq 2$. Write $x = x_1 + px_2$ where x_1 runs modulo p , and x_2 runs modulo p^{c-1} ; then the sum over x_2 vanishes unless $y \equiv -m_1 \pmod{p^{c-1}}$, so we have

$$a(m; p^c, \pm p) = p^c \sum_{x_1 \pmod{p}} \sum_{y \equiv -m_1 \pmod{p^{c-1}}}^* \left(\frac{x_1}{p}\right) e\left(\frac{x_1(y+m_1)}{p^c}\right) e\left(\frac{\bar{y}}{p^c}\right). \quad (10.38)$$

Now write $y \equiv -m_1 + vp^{c-1}$ where v runs modulo p . The sum over x_1 becomes a Gauss sum. Noting that $\bar{y} \equiv -\bar{m}_1 - v\bar{m}_1^2 p^{c-1}$, we have

$$a(m; p^c, \pm p) = p^{c+\frac{1}{2}} \epsilon_p \sum_{v \pmod{p}} \left(\frac{v}{p}\right) e\left(\frac{-\bar{m}_1 - v\bar{m}_1^2 p^{c-1}}{p^c}\right) = p^{c+1} e\left(\frac{-\bar{m}_1}{p^c}\right). \quad (10.39)$$

Having considered all the cases, this completes the proof of (10.30).

Finally, we show $a(0; c, D) = 0$. To see this, it suffices to note from the above calculations that $a(0; q^c, 1) = 0$ for $c \geq 1$. Since $(q, D) = 1$, we have that $a(0; c, D)$ is divisible by $a(0; q^c, 1)$ for some such $c \geq 1$. \square

Proof of Lemma 10.6. We begin with the important special case $T \leq X^\varepsilon$. The trivial bound shows $r(m; c, D) \ll (T+1)MN \ll NX^\varepsilon$. For $m \neq 0$, integration by parts twice shows that

$$r(m; c, D) = \frac{-(cD)^2}{(2\pi m)^2} \int_0^\infty \frac{d^2}{dx^2} \left[w(x) H\left(\frac{4\pi\sqrt{x}}{c}\right) \right] e\left(\frac{-mx}{c|D|}\right) dx \ll \left(\frac{|D|}{|m|\sqrt{N}}\right)^2 N(T+1)MX^\varepsilon, \quad (10.40)$$

which when combined with the trivial bound gives (10.31).

Now suppose $T \geq X^\varepsilon$ and hence $M \geq X^\varepsilon$, adjusting the value of ε as necessary. For convenience, note that integrating by parts one more time in (10.40) shows $r(m; c, D) \ll \left(\frac{|D|}{|m|\sqrt{N}}\right)^3 N(T+1)MX^\varepsilon$ which is satisfactory for (10.31) provided $|m| \gg X^{100}$, say. Now assume $|m| \ll X^{100}$. Inserting (10.18) into the definition of $r(m; c, D)$, writing $2 \cos(y) = e^{iy} + e^{-iy}$, and changing variables $x \rightarrow Nx$, we have

$$r(m; c, D) = \frac{N(T+1)}{2} \sum_{\delta_1, \delta_2 \in \{\pm\}} \int_{-\infty}^\infty \int_0^\infty w_N(x) u_{\delta_1}(v) e\left(-\frac{mN}{c|D|}x + \delta_1 \frac{Tv}{M} + 2\delta_2 \sqrt{x} \frac{\sqrt{N}}{c} \cosh\left(\frac{\pi v}{M}\right)\right) dx dv, \quad (10.41)$$

where $w_N(x) = w(Nx)$ so that $w_N^{(j)}(x) \ll_j 1$.

By the rapid decay of u_δ , and the fact that m is polynomially bounded by X , we may truncate the v -integral at M^ε since this gives an acceptable error term. Now consider the inner x -integral. With

$$F_1(x) = -\frac{mN}{c|D|}x + \delta_2 2\sqrt{x} \frac{\sqrt{N}}{c} \cosh(\pi v/M), \quad (10.42)$$

we have

$$F_1'(x) = -\frac{mN}{c|D|} + \delta_2 \frac{\sqrt{N}}{c\sqrt{x}} \cosh(\pi v/M), \quad (10.43)$$

and for $j \geq 2$ we have

$$F_1^{(j)}(x) \asymp \frac{\sqrt{N}}{c} \cosh(\pi v/M) \asymp \frac{\sqrt{N}}{c}, \quad (10.44)$$

for all x in the support of w_N . An easy application of Lemma 8.1 of [BKY] shows that the x -integral is very small unless

$$\left| \frac{mN}{cD} \right| \asymp \frac{\sqrt{N}}{c}. \quad (10.45)$$

Precisely, if (10.45) does not hold, then

$$\int_{-\infty}^{\infty} w_N(x) e(F_1(x)) dx \ll_A \left(\left| \frac{mN}{cD} \right| + \frac{\sqrt{N}}{c} \right)^{-A}. \quad (10.46)$$

Since $\sqrt{N}/c \geq X^\varepsilon$ (adjusting ε as necessary), the contribution to $r(m; c, D)$ for parameters where (10.45) does not hold is satisfactory for (10.31). Now assume (10.45) holds. Before refining the x -integral, we turn to the v -integral, which takes the form

$$\int_{|v| \leq M^\varepsilon} u_\pm(v) e(F_2(v)) dv, \quad (10.47)$$

where

$$F_2(v) = \delta_1 \frac{Tv}{M} + 2\delta_2 \sqrt{x} \frac{\sqrt{N}}{c} \cosh\left(\frac{\pi v}{M}\right). \quad (10.48)$$

We calculate

$$F_2'(v) = \delta_1 \frac{T}{M} + 2\delta_2 \sqrt{x} \frac{\sqrt{N}}{c} \frac{\pi}{M} \sinh\left(\frac{\pi v}{M}\right), \quad (10.49)$$

and for $j \geq 2$,

$$F_2^{(j)}(v) \ll \frac{\sqrt{N}}{c} \frac{1}{M^j}. \quad (10.50)$$

Similarly,

$$\frac{d^j}{dv^j} w_0\left(\frac{x}{Z} - \frac{D^2 \cosh^2(\pi v/M)}{m^2 N Z}\right) u_\pm(v) \quad (10.51)$$

If $c \geq \frac{\sqrt{N}}{MT} X^\varepsilon$, then $F_2' \gg \frac{T}{M}$ and another easy application of Lemma 8.1 of [BKY] shows that the v -integral is very small. Now suppose

$$c \leq \frac{\sqrt{N}}{MT} X^\varepsilon. \quad (10.52)$$

Now we return to the x -integral. With an appropriate choice of δ_2 , a stationary point exists inside the range of integration at

$$x_0 = \frac{D^2 \cosh^2(\pi v/M)}{m^2 N} \sim \frac{D^2}{m^2 N}. \quad (10.53)$$

For the other choice of sign of δ_2 the first derivative is large and the integral is very small. The proof of Proposition 8.2 of [BKY] shows that with $Z = Y^{-1/2+\varepsilon}$, $Y = \sqrt{N}/c$, we have

$$\int_{-\infty}^{\infty} w_N(x) e(F_1(x)) dx = \int_{-\infty}^{\infty} w_N(x) w_0\left(\frac{x-x_0}{Z}\right) e(F_1(x)) dx + O(X^{-A}), \quad (10.54)$$

where w_0 is any fixed, compactly-supported function that is identically 1 in a neighborhood of 0. The point is that on the complement of $|x-x_0| \leq Z$ the first derivative of F_1 is large

enough that repeated integration by parts shows that the complementary integral is very small. Thus we have

$$|r(m; c, D)| \ll NT \int_{-\infty}^{\infty} \left| \int_{|v| \leq M^\varepsilon} K(v) e(F_2(v)) dv \right| dx + O(X^{-100}), \quad (10.55)$$

for some choice of δ_i 's, where

$$K(v) = w_0 \left(\frac{x}{Z} - \frac{D^2 \cosh^2(\pi v/M)}{m^2 N Z} \right) u_\pm(v). \quad (10.56)$$

Note $K^{(j)}(v) \ll Z^{-1} M^{-(1-\varepsilon)j}$. Proposition 8.2 of [BKY] again shows that the inner v -integral in (10.55) is bounded by $M^{1+\varepsilon}/\sqrt{Y}$, giving now

$$r(m; c, D) \ll NT \frac{M}{Y} X^\varepsilon = N \frac{cMT}{\sqrt{N}} X^\varepsilon \ll NX^\varepsilon, \quad (10.57)$$

using (10.52).

Finally, we remark that if $T \gg X^\varepsilon$ then the condition (10.45) is incompatible with $m = 0$, so $r(0; c, D) = O(X^{-100})$. \square

11. EXTENDING THE KOHNEN AND SENGUPTA RESULT

Here we quickly sketch the proof of (1.11). Formula (8) of [KoSe] gives that the left hand side is

$$\ll k^{1+\varepsilon} \left(1 + \left| \sum_{c \geq 1} \frac{1}{c} H_c(|D|, |D|) J_{k-\frac{1}{2}}(\pi|D|/c) \right| \right), \quad (11.1)$$

where $H_c(m, n)$ is a generalized Kloosterman sum corresponding to the half-integral weight multiplier system. By calculations in Section 3 of [I1], we have the bound

$$|H_c(m, n)| \leq 2^{\nu+2} |S_\chi(m', n'; l)|, \quad (11.2)$$

where $c = 2^\nu l$ with l odd, $m' = \bar{2}^\nu m$, $n' = \bar{2}^\nu n$, and $S_\chi(m, n; l)$ is the usual Salié sum. Then by Lemmas 3 and 4 of [I1], we have $|S_\chi(n, n; l)| \leq d(l) l^{1/2} (n, l)^{1/2}$. Since $k \rightarrow \infty$, we can truncate the sum over c at $c \leq 100|D|/k$ since otherwise the Bessel function is exponentially small. We have for $x \gg \nu$ that (see (2.11') of [ILS])

$$J_\nu(x) \ll \nu^{-1/4} (|x - \nu| + \nu^{1/3})^{-1/4}.$$

Combining these bounds and summing trivially over c , we obtain (1.11), as desired.

On the side, we remark that the general approach used to prove Theorem 1.5 can be used to prove (1.11) also.

REFERENCES

- [Bi] A. Biró, *Cycle integrals of Maass forms of weight 0 and Fourier coefficients of Maass forms of weight 1/2*. Acta Arith. 94 (2000), no. 2, 103–152.
54 (2004), 831–847.
- [Bl] V. Blomer, *On the 4-norm of an automorphic form*, Journal of the European Mathematical Society, to appear. arXiv:1110.4717
- [BH] V. Blomer and G. Harcos, *Hybrid bounds for twisted L-functions*. J. Reine Angew. Math. 621 (2008), 53–79.
- [BKY] V. Blomer, R. Khan, and M. Young, *Mass distribution of holomorphic cusp forms*. Preprint, 2012, <http://arxiv.org/abs/1203.2573>
- [CI] B. Conrey and H. Iwaniec, *The cubic moment of central values of automorphic L-functions*. Ann. of Math. 151 (2000), 1175–1216.

- [D] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*. Invent. Math. 92 (1988), no. 1, 73–90.
- [DFI] W. Duke, J. Friedlander, and H. Iwaniec, *Class group L -functions*. Duke Math. J. 79 (1995), no. 1, 1–56.
- [FW] B. Feigon and D. Whitehouse, *Averages of central L -values of Hilbert modular forms with an application to subconvexity*. Duke Math. J. **149** (2009), 347–410.
- [GJ] S. Gelbart and H. Jacquet, *A relation between automorphic representations of $GL(2)$ and $GL(3)$* . Ann. Sci. École Norm. Sup. **11** (1978), 471–542.
- [GR] I.S. Gradshteyn, and I.M. Ryzhik, *Table of Integrals, Series, and Products*. Translated from the Russian. Sixth edition. Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger. Academic Press, Inc., San Diego, CA, 2000.
- [GZ] B. Gross and D. Zagier, *Heegner points and derivatives of L -series*. Invent. Math. **84** (1986), 225–320.
- [HaMi] G. Harcos and P. Michel, *The subconvexity problem for Rankin-Selberg L -functions and equidistribution of Heegner points. II*. Invent. Math. **163** (2006), 581–655.
- [HaTe] G. Harcos and N. Templier, *On the sup-norm of Maass cusp forms of large level. III*. Preprint.
- [H-B1] D.R. Heath-Brown, *Hybrid bounds for Dirichlet L -functions*. Invent. Math. 47 (1978), no. 2, 149–170.
- [H-B2] D.R. Heath-Brown, *Hybrid bounds for Dirichlet L -functions. II*. Quart. J. Math. Oxford Ser. (2) 31 (1980), no. 122, 157–167.
- [HoMu] R. Holowinsky and R. Munshi, *Level Aspect Subconvexity For Rankin-Selberg L -functions*. arXiv:1203.1300
- [HoTe] R. Holowinsky and N. Templier, *First moment of Rankin-Selberg central L -values and subconvexity in the level aspect*. <http://arxiv.org/abs/1207.3421>
- [HW] M. N. Huxley and N. Watt, *Hybrid bounds for Dirichlet’s L -function*. Math. Proc. Cambridge Philos. Soc. 129 (2000), no. 3, 385–415.
- [I1] H. Iwaniec, *Fourier coefficients of modular forms of half-integral weight*. Invent. Math. 87 (1987), no. 2, 385–401.
- [I2] H. Iwaniec, *Topics in Classical Automorphic Forms*. Grad. Stud. Math., vol 17, Amer. Math. Soc., 1997.
- [IK] H. Iwaniec and E. Kowalski, *Analytic number theory*. American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004. xii+615 pp.
- [ILS] H. Iwaniec, W. Luo and P. Sarnak, *Low lying zeros of families of L -functions*. Publ. Math. Inst. Hautes Études Sci. no. 91, 55-131 (2001).
- [KaSa] S. Katok and P. Sarnak, *Heegner points, cycles and Maass forms*. Israel J. Math. 84 (1993), no. 1-2, 193–227.
- [KoSe] W. Kohnen and J. Sengupta, *On quadratic character twists of Hecke L -functions attached to cusp forms of varying weights at the central point*. Acta Arith. 99 (2001), no. 1, 61–66.
- [KMV] E. Kowalski, P. Michel, J. VanderKam, *Rankin-Selberg L -functions in the level aspect*. Duke Math. J. **114** (2002), 123–191.
- [L] Xiannan Li, *Upper bounds on L -functions at the edge of the critical strip*. IMRN (2010), 727–755.
- [LY] Xiaoqing Li and M. Young, *The L^2 restriction norm of a GL_3 Maass form*. Compositio Math. **148** (2012), 675–717.
- [Mi] P. Michel, *The subconvexity problem for Rankin-Selberg L -functions and equidistribution of Heegner points*. Ann. of Math. **160** (2004), 185–236.
- [MR] P. Michel and D. Ramakrishnan, *Consequences of the Gross/Zagier formulae: Stability of average L -values, subconvexity, and non-vanishing mod p* . arXiv:0709.4668
- [MV1] P. Michel and A. Venkatesh, *Equidistribution, L -functions and ergodic theory: on some problems of Yu. Linnik*. International Congress of Mathematicians. Vol. II, 421–457, Eur. Math. Soc., Zürich, 2006.
- [MV2] P. Michel and A. Venkatesh, *Heegner points and non-vanishing of Rankin/Selberg L -functions*. Analytic number theory, 169–183, Clay Math. Proc., 7, Amer. Math. Soc., Providence, RI, 2007.
- [MV3] P. Michel and A. Venkatesh, *The subconvexity problem for GL_2* . Publ. Math. Inst. Hautes Études Sci. No. 111 (2010), 171–271.
- [Mu] R. Munshi, *On a hybrid bound for twisted L -values*. Arch. Math. (Basel) 96 (2011), no. 3, 235–245.
- [N1] P. Nelson, *Equidistribution of cusp forms in the level aspect*. Duke Math. J. 160 (2011), no. 3, 467–501.
- [N2] P. Nelson, *Stable averages of central values of Rankin-Selberg L -functions: some new variants*. arXiv:1202.6313

- [S] W. Schmidt, *Equations over finite fields. An elementary approach*. Lecture Notes in Mathematics, Vol. 536. Springer-Verlag, Berlin-New York, 1976. ix+276 pp.
- [Te] N. Templier, *A nonsplit sum of coefficients of modular forms*. Duke Math. J. 157 (2011), no. 1, 109–165.
- [Ti] E. C. Titchmarsh, *The theory of the Riemann zeta-function*. Second edition. Edited and with a preface by D. R. Heath-Brown. The Clarendon Press, Oxford University Press, New York, 1986. x+412 pp.
- [Wal] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier*. J. Math. Pures Appl. (9) 60 (1981), no. 4, 375–484.
- [Wats] T. Watson, *Rankin triple products and quantum chaos*, Annals of Mathematics, to appear.
- [Watt] N. Watt, *On the mean squared modulus of a Dirichlet L-function over a short segment of the critical line*. Acta Arith. 111 (2004), no. 4, 307–403.
- [Z1] S. W. Zhang, *Gross–Zagier formula for GL_2* . Asian J. Math. 5 (2001), 183–290.
- [Z2] S. W. Zhang, *Equidistribution of CM-points on quaternion Shimura varieties*. Int. Math. Res. Not. 2005, no. 59, 3657–3689.

DEPARTMENT OF MATHEMATICS, MAILSTOP 3368, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77843-3368

E-mail address: `scliu@math.tamu.edu`

E-mail address: `masri@math.tamu.edu`

E-mail address: `myoung@math.tamu.edu`